



## Linguística de *corpus* e fraseologias em séries televisivas: uma experiência que deu certo

Márcio Issamu Yamamoto<sup>1</sup>  
Joel Victor Reis Lisboa<sup>2</sup>

**Resumo.** Neste artigo objetivamos trabalhar alguns conceitos de fraseologias, dentre eles o de Tagnin (2005), Barbosa (2012), Dobrovol'skij (2012) e Muryn et al. (2013), bem como os conceitos de metáfora de Lakoff e Johnson (2003). Também abordaremos aspectos formais, semânticos e pragmáticos das fraseologias. Ademais, classificaremos as fraseologias das pesquisas terminográficas e fraseológicas de Iniciação Científica da Universidade Federal de Uberlândia (UFU) e da Prática como Componente Curricular da Universidade Federal de Jataí (UFJ) com base na fundamentação teórica. As fraseologias bilíngues, inglês-português, foram extraídas de *corpora* de legendas de séries televisivas, como *Bones*, *Game of Thrones* e *Star Trek*, sendo que neste artigo apresentamos 30 delas. A metodologia adotada foi a Linguística de *Corpus* (LC) e o *WordSmith Tools* 6.0 foi o programa de análise lexical utilizado. Após a escolha da série, partimos para a compilação das legendas em português e inglês na internet. Em seguida, fizemos as listas de palavras e escolhemos os termos e verbos que permitiram a seleção e a posterior análise das fraseologias dos seriados por meio do Concordanciador. A maior parte dos termos analisados foi registrada no VoTec (FROMM, 2007), plataforma de pesquisa terminográfica online de acesso público. Os alunos que desenvolveram essas pesquisas são graduandos do curso de Letras Inglês, e o público-alvo são tradutores, especialistas da linguagem, professores, alunos e fãs dos seriados. Algumas vantagens que estas pesquisas trouxeram foram: (1) análise do uso real da língua inglesa por falantes nativos em contextos comunicativos autênticos; (2) acesso dos graduandos à LC e ao fazer fraseológico/terminográfico; e (3) capacitação para desenvolvimento de material didático para aulas de língua inglesa.

**Palavras-chave:** Fraseologia bilíngue. Linguística de *Corpus*. Seriados televisivos.

### *Corpus* linguistics and phraseologies on tv series: a successful experience

**Abstract.** In this paper we aim to work on some phraseology concepts, including the ones proposed by Tagnin (2005), Barbosa (2012), Dobrovol'skij (2012), Muryn et al. (2013), as well as on metaphor concepts by Lakoff and Johnson (2003). Furthermore, we will discuss formal, semantic and pragmatic aspects of phraseologies and use the theoretical framework to classify the results of phraseological and terminographical studies from the Federal University of Uberlândia (UFU) Scientific Initiation and the Federal University of Jataí (UFJ) Prática como Componente Curricular. The English-Portuguese phraseologies were extracted from *corpora* of TV series subtitles, such as *Bones*, *Game of Thrones* and *Star Trek*, 30 being selected for this paper discussion. The methodology adopted was *Corpus* Linguistics (CL) and *WordSmith Tools* 6.0 was the lexical analysis program used. We started by selecting the TV series, then Portuguese and English subtitles were compiled from the internet. Afterwards, word lists were made; we chose terms and verbs that allowed the selection and analysis of phraseologies from TV series by using the Concord tool and most of the analyzed terms were registered on VoTec (FROMM, 2007), an online platform for terminological and terminographical research open to public access. The students that carried out these researches are English Language and Literature undergraduates, and the

<sup>1</sup> Doutorando na Universidade Federal de Uberlândia. [issamu2009@gmail.com](mailto:issamu2009@gmail.com)

<sup>2</sup> Mestrando na Universidade Federal de Uberlândia. [joelvictorlisboa@gmail.com](mailto:joelvictorlisboa@gmail.com)



target audience are translators, language specialists, teachers, students and TV series fans. Some advantages of these researches were: (1) analysis of the real use of English language by native speakers in authentic communicative contexts; (2) access of undergraduates to CL and phraseological/terminographical studies; and (3) training to develop English language teaching material.

**Keywords:** Bilingual phraseology. *Corpus* Linguistics. TV series.

## 1 INTRODUCTION

The aim of this paper is to work on various concepts of phraseology and metaphor, and the theoretical framework will be based on Lakoff and Johnson (2003), Tagnin (2005), Dobrovol'skij (2012), Barbosa (2012) and Muryn et al. (2013). Furthermore, we will seek to approach formal, semantic and pragmatic aspects of phraseologies and use the theoretical framework to classify the results of Portuguese-English phraseological and terminographical researches.

The examples of phraseologies or phraseological units (PU) are derived from English Language and Literature students' academic work from the Federal University of Uberlândia (UFU) and the Federal University of Jataí (UFJ). For the phraseologies and terms selection, *corpora* from TV series subtitles were compiled and analyzed based on *Corpus* Linguistics (CL), which was adopted as a methodology and approach for this research. In this paper, we will present 30 phraseologies from TV series such as *Bones*, *Game of Thrones* and *Star Trek*, whose subtitles were treated by the lexical analysis program WordSmith Tools 6.0 (SCOTT, 2012), hereafter WST.

The results of the Scientific Initiation (SI) researches and part of the results of Prática como Componente Curricular (PCC)<sup>3</sup> researches were made available on VoTec<sup>4</sup> platform (FROMM, 2007), in which the entries, definition, examples and other constituent elements of the microstructure were made available. The partial results of other PCC works are included in this paper.

## 2 THEORETICAL FRAMEWORK

<sup>3</sup> Prática como Componente Curricular: the aim of this student formation is to decrease the existing gap between theoretical knowledge and its praxis, consisting of a 50-hour project by term. All in all, students will have performed a 400-hour PCC before getting their degree (BA in English Language and Literature). More details available at: < [https://ingles.jatai.ufg.br/up/702/o/PPC\\_Letras.pdf?1431712586](https://ingles.jatai.ufg.br/up/702/o/PPC_Letras.pdf?1431712586) > (in Portuguese).

<sup>4</sup> Platform available at: < <http://ic.votec.ileel.ufu.br/> >.



In this first section, we will discuss theoretical aspects of phraseological studies and present some expressions to exemplify the presented concepts.

## 2.1 Phraseology: concepts

Nowadays, we can notice that phraseology bears various concepts in the scholar environment. Besides that, Phraseology<sup>5</sup> is still a field underway to its consolidation in the Brazilian academic context (BEVILACQUA, 1996; BARBOSA, 2012).

First, of all, Phraseology can be conceived as a (1) subdiscipline of Lexicology (KLARE, 1986), (2) a linguistic discipline or (3) a branch/subdiscipline of Linguistics dealing with word combinations, that be locutions, idioms, fixed or institutionalized expressions, complex lexias, paroemias (proverbs) and sayings (BEVILACQUA, 1996; FLEISCHER, 1997; BARBOSA, 2012; DOBROVOL'SKIJ, 2012; CHACOTO, 2012; ALVAREZ, 2012, 2014).

Secondly, phraseology is a term that encompasses expressions or phrasemes present in natural languages, which bring with them distinctive marks such as idiomaticity (with greater or lesser degree), conventionality (institutionalized by use), structural fixedness or rigidity of lexical items (BALLY, 1909; FLEISCHER, 1997; ÁLVAREZ, 2012; BARBOSA, 2012; DOBROVOL'SKIJ, 2012; CHACOTO, 2012).

The contribution of Lakoff and Johnson (2003) to phraseology concept comprehension is due to their theory about metaphors and their relation to phrasemes idiomaticity. The authors state, "We see metaphor as essential to human understanding and as a mechanism for creating new meaning and new realities in our lives" (LAKOFF; JOHNSON, 2003, p. 196). Metaphors creative power and their ability to resurrect words from general lexicon in a different meaning is what explains idiomaticity in phraseologies. Barbosa (2012) explains that this process occurs when we analyze the effect of phraseologies from the point of view of norms, which marks a universe of specific discourse.

---

<sup>5</sup> Phraseology here is capitalized to indicate its conception as a branch of some Linguistics discipline or as a discipline. In the future, it may be that Phraseology will become a specific subdiscipline of Linguistics, depending on its development from the theoretical-methodological point of view, as it happened with Lexicology and Terminology.



According to Alvarez (2014, p. 37), "to become an idiomatic expression it undergoes a process of metaphorization"<sup>6</sup>, that is where we find the interrelationship between phraseologies and metaphors. Then, Lakoff and Johnson (2003) explain that our conceptual system is metaphorical, systematic and cultural. This system governs our daily lives in every way, in the structuring of language, in the perception of reality, in the development of interpersonal relationships and in how we relate to the world. Alvarez (2014, p. 22) ratifies this concept by stating that "recent research points to the fact that in our day-to-day interactions we use at least four metaphors per minute". Consequently, we understand phraseologies are systematically linked to metaphorical concepts that help us understand the symbolic nature of peoples' social, culture and language activities (DOBROVOL'SKIJ; PIIRAINEN, 2005; ÁLVAREZ, 2012; BARBOSA, 2012).

From a cultural point of view, both Lakoff and Johnson (2003) and Alvarez (2012) relate cultural identity to verbal expressions of their speakers and societies, whether by metaphors or phraseologies. At this point of our discussion, we can consider both as synonymous, since they are linked by the same conceptual network that gives form to language, of metaphorical nature, with a greater or lesser degree of literality. Lakoff and Johnson (2003) exemplify this concept with expressions such as **time is money** and its settings/expressions: **gain/lose money** and **spend/waste money**, conceivable in some cultures, but not in all cultures worldwide. Another example brought is the spatial concept of future: it is conceived as being ahead in some cultures, whereas in others, it is spatially situated behind and not forward (LAKOFF; JOHNSON, 2003, p. 8-10, 16).

This intimate constitutive relationship between cultural identity and linguistic expressions becomes evident when we analyze Western countries TV series. In these countries, of mostly Judeo-Christian cultural identity, we have identified, in the great majority, phraseologies distinct from those that exist in Eastern countries. For example, in Hell on

---

<sup>6</sup> Dias (2010 apud ALVAREZ, 2014, p. 37) defines metaphorization as a "process based on the possibility of establishing a similarity relation between two objects, that is, correspondence between distinct units, due to the existence of common semes."



Wheels<sup>7</sup>, the phraseology **for God's sake** appears 6 times, while Mr. Sunshine<sup>8</sup> this same expression does not appear even once.

## 2.2 Phraseologies: structure and semantics

Firstly, from a structuralist perspective, phraseologies are syntactic structures composed of relatively fixed phrases, with restrictions in the combinations, and may present some flexibility, that is, the position of the constituent elements within the syntagma can vary (MURYN et al., 2013).

Barbosa (2012, p. 249) defines phraseologies as "a class of syntactic and semantic equivalence; class of elements that are not identical, but that can be grouped according to some criterion". The criterion that groups them is the degree of integration between them. For example, in the English expression **to get away with murder**, to leave unharmed of an accusation or to leave a situation without being punished, the combination **to get away with** cannot be modified as to its syntax, according to the Corpus of Contemporary American English (COCA)<sup>9</sup>. As examples we have: "he was now the guy who helped him get away with murder and he was destroyed by everybody". The possible change is made possible if the crime type is specified before, then the pronoun **it** can be used to retake the idea *a posteriori*: "that folks do not come into my town to steal and murder and get away with it." When it comes to phraseology extensions, they go from two lexical items, until a complete phrase, including proverbs and sayings, also called phrasemes.

Second, from the semantic point of view, Dobrovolskij (2012) explains that phraseologies have a conceptual structure underlying their semantic aspect, consisting of two levels (macrocomponents): (1) the figurative or metaphorical meaning and (2) the literal meaning of the lexical constituents. In order to understand them, depending on their greater or lesser degree of idiomaticity, it is necessary to stick to the first level, since the metaphorical

---

<sup>7</sup> An American/Canadian Series that portrays the Pilgrims' arrival from Europe to the United States, when of the First Transcontinental Railroad construction (East-West), in the late nineteenth century. More info at: <<https://www.amc.com/shows/hell-on-wheels>>.

<sup>8</sup> An original Netflix Korean series depicting the occupation of Korea by Japan and the United States in the late 19th and early 20th centuries. The characters are Japanese, Koreans and Americans living in a Buddhist context. More info at: <<https://www.netflix.com/br/title/80991107>>.

<sup>9</sup> *Corpus* available at: <https://corpus.byu.edu/coca/>. Access on Sept. 10th, 2018. We brought up two examples, even though there are various available on the platform.



meaning of some phraseologies is dissociated from the mental image or literal meaning of each constituent.

In relation to these two macrocomponents, the metaphorical and the literal meaning, both serve as a starting point and as a conceptual foundation for semantic reinterpretation. Consequently, there are idiomatic expressions in different languages that share the same mental image, but having totally different meanings (ÁLVAREZ, 2012). As an example, we have the image of **kick the bucket** in English and *chutar o balde* in Portuguese. The image is the same, but meanings are totally different. Whereas in English the meaning is “to die”, in Portuguese it means “to give up something” or “to lose control”. This semantic aspect will be detailed in item 2.4, on the idiomaticity of phraseologies. Next, we will present some types of phraseologies according to Tagnin (2005) and Dobrovolskij (2012).

### 2.2.1 Colligation and collocations

Tagnin (2005), when dealing with conventionality<sup>10</sup> and sentence structure, proposes classifications as (1) **colligation**: word of content (WC)<sup>11</sup> + grammatical word, and (2) **collocation**: WC + WC.

Firstly, the author describes the types of colligations: (1a) **verbal regency colligation** (WC + grammatical word<sup>12</sup>: devote to/*dedicar-se a*, aptitude for/*aptidão para*, dressed in/*vestido(a) de*; instead of/*em vez de*, phrasal verbs<sup>13</sup>); (1b) **colligations of prepositions** (Prep.+ Noun Phrase (NP): at your discretion/*a seu critério*; Prep.+ NP + Prep.: at a cost of/*a custo de*).

Next, Tagnin brings the classification and examples of collocations: (2a) **adjective collocation** (Adj. + WC: silent movie/*cinema mudo*, brown sugar/*açúcar mascavo*); (2b) **noun collocation** (coarse salt/*sal grosso*, confectioner’s sugar/*açúcar de confeitiro*, skinflint/*pão-*

<sup>10</sup> Conventionality, according to Tagnin (2005), encompasses all that is of use or custom; what is consolidated by use or practice.

<sup>11</sup> What Tagnin (2005) names word of content is the same as what Aničkov (1937, p. 105 apud DOBROVOL’SKIJ, 2012, p. 18) names autosemantic word, when defining colligation as “word combinations consisting of one autosemantic word and one synsemantic word (to look at home)”. What the author names synsemantic word would be what is known in Lexicology as a grammatical word.

<sup>12</sup> Also called function word.

<sup>13</sup> The author specifies the various possible structures of phrasal verbs and describes in detail the fixedness or rigidity aspects of this type of phrasing typical of English and German (TAGNIN, 2005, p. 33-35). We recommend this reading for those interested in delving into the topic.



*duro*, tree of knowledge/*árvore do conhecimento*); (2c) **verb collocation** (V + Object: bring suit/*abrir processo*; V + prep. + N<sup>14</sup>: put to vote/*pôr em votação*); (2d) **adverb collocation** (lavishly illustrated/*fartamente ilustrado*; love blindly/*amar cegamente*)<sup>1516</sup>.

### 2.2.2 Idiomatic expressions

According to Dobrovol'skij (2012), idiomatic expressions can be **phrasemes** or **restricted collocations**. As **phrasemes** we have kick the bucket/*bater as botas* and spill the beans/*dar com a língua nos dentes*. Besides this hypernym, the author points out its subdivisions: the **binomials** (by and large/*em geral*), the **simile** (sleep like a log/*dormir como uma pedra*) and **speech formulae** (be that as it may be/*seja como for*). Another subgroup would be the **restricted collocations** (strong tea/*chá forte*), and those based on verbs (take measures/*tomar medidas*). In this type of phraseology, we observe that both Tagnin (2005) and Dobrovol'skij (2012) address the same structures, name and exemplify them.

### 2.2.3 Paroemias: Proverbs and Sayings

Besides the phraseology types already discussed on this paper, there are the most complex phrasemes in terms of their structural, lexical-semantic and pragmatic aspects: **proverbs** and **sayings**, also called **paroemias**. Language scientists agree that differentiating them is not an easy task, since each theoretical current defines them in a different way. Therefore, we will bring some definitions proposed by Dobrovol'skij (2012).

Dobrovol'skij (2012, p. 23) states that proverbs are small texts of ethnocultural origin, ethnographic or folkloric as **the first time is always the hardest** and **early to bed and early to rise (makes a man healthy, wealthy, and wise)**. As traits of identity, they have the following characteristics:

<sup>14</sup> V+prep+N: Verb + preposition + noun

<sup>15</sup> In the first example of 2d, the adverb modifies the adjective; in the second, the adverb modifies the verb.

<sup>16</sup> In addition to colligations and collocations, used in this paper, Tagnin (2005, p. 47-61) also names specific expressions of units, collective, binomials and consecrated agrammatic structures (syntactically unpredictable structures, syntactically petrified structures, unpredictable syntactic block) and conventional expressions. This second part of the structural and conventional aspect of the phraseological units will not be part of the application in this paper.

1. Their semantic content reveals a universal truth: **slow and steady wins the race; trouble comes to him who seeks it.**
2. Sharing the features of opinions, recommendations on how to behave or not in certain situations: **don't look a gift horse in the mouth; all is not gold that glitters, look before you leap.**
3. Proverbs of advice are based on worldly wisdom shared by a specific linguistic community, that is, they are dependent on a culture and its specificity. *Hito no kuchi ni to wa tateranu* (one can't control what is said by others – Japanese proverb) and **little strokes fell great oaks.**

In contrast to proverbs, in conventional sayings and expressions there is no trace of universality, recommendation, or explanation. In proverbs, these characteristics are marked by the imperative verbal forms (one does not look ..., one does/shall not ..., do as ...), adverbs (not everything...), and they are meaningful despite contexts.

#### 2.2.4 Idiomaticity

The distinctive feature of some phraseologies in relation to common lexical combinations is the idiomaticity. It is a feature that makes the combination of lexical items, constituent of phraseologies, not retain its original meaning, but acquire a metaphorical meaning. Álvarez (2012) based on Corpas Pastor (1999) defines idiomaticity as follows,

es decir que el significado global de una unidad fraseologica no es deducible del significado aislado de cada uno de sus elementos constitutivos, pues presentan especialización o lexicalización semántica, por la cual su significado denotativo es figurativo o traslaticio, debido, entre otros, a procesos metafóricos y/o metonímicos (ÁLVAREZ, 2012, p. 172).<sup>17</sup>

Alvarez (2012) points out the aspect of some phraseologies being idiomatic expressions, the fruit of a metaphorical conceptual network or the fruit of metaphors, agreeing with Lakoff and Johnson (2003). Lewis (1999) explains the idiomaticity by fixed expressions concept, explaining that in **medium Coke** combination, the lexical constituents retain their literal

---

<sup>17</sup> “that is to say that the global meaning of a phraseological unit is not deductible from the isolated meaning of each one of its constituent elements, since they present specialization or semantic lexicalization, by which its denotative meaning is figurative or metaphorical, due, among others, to metaphorical processes and/or metonymics.” (ÁLVAREZ, 2012, p. 172, own translation).



meaning, whereas in the expression **twin rooms**, there is a partial semantic correspondence with **twins** as a noun. That is, when we use the adjective **twin** in **twin rooms** as a collocation, the meaning of **twin** combines with **rooms** and produces a partial semantic correspondence of the term **twins** literal meaning, resulting in a phenomenon of partial similarity of meaning or semantic displacement.

Similar to what was proposed in the previous paragraph, Dobrovolskij (2012) mentions Vinogradov (1947 apud DOBROVOL'SKIJ, 2012) and three phraseological classifications based on sign motivation, such as: (1) **frozen idioms**: phraseology in which the meaning is different from the individual meaning of each constituent, as in **to show somebody Kuzma's mother** (to show somebody who is the boss – Russian proverb); (2) **motivated idioms**: when there is a partial resemblance between the phraseology constituents and its meaning, as in **one is very healthy-looking, with good facial color: the (very) picture of health** (Russian proverb); (3) **restricted collocations**, such as **to take a snapshot, to take a photo**, in which the meaning can be derived from their lexical items. In this case, a component, such as **to take**, can be marked by its opacity<sup>18</sup>, and in an isolated analysis it would be difficult to define its meaning.

A structural and semantic opacity example is the Portuguese institutionalized expression/phraseology *Pois não*, which causes estrangement to Spanish and French speakers, learners of Portuguese as a Foreign Language. The point is, if this expression bears an affirmative meaning, why should there be a **no** in it? This example portrays well the cultural and pragmatic aspect of phraseologies.

### 2.3 Phraseology: pragmatic aspects

In pragmatic scope, phraseologies are used in contexts in which there are specific interpersonal relations, with illocutionary objectives and determined situations and interlocutors. The sociological and psycholinguistic approach of language conceives culture as a set of shared beliefs, common places, stereotypes and pragmatic restrictions. Therefore, culture governs the use of phraseologies. That is, language shaped by cultural molds; or culture

---

<sup>18</sup> This type of lexical item is classified by Lewis (1999) as delexicalized verbs, that is, verbs with an extensive range of meanings, so that it is almost impossible to define its meaning when isolated from its constitutive expressions or context.



embodied in the language as it was pointed out in the usage of **bucket** in previous expressions (ALVAREZ, 2012; MURYN et al., 2013).

### 3 METHODOLOGY

In this section, we will briefly present CL and the WST, and how they were used by undergraduates to extract phraseologies and terms in SI and PCC researches of UFU and UFJ English Language and Literature courses. The other phraseologies were selected from VoTec platform and from the *corpora* of TV series subtitles.

#### 3.1 *Corpus Linguistics*

CL is an interdisciplinary branch of language studies (BERBER SARDINHA, 2000; BERBER SARDINHA et al., 2017) which conceives language as a probabilistic system of choices, in which not everything that is possible is likely to actually happen in terms of probability (HALLIDAY, 1991). CL works on *corpora* compilation and analysis, those being collections of authentic oral or written texts, compiled and carefully organized for the purpose of language studies (BERBER SARDINHA, 2003; GARCIA, 2008). In the SI and PCC researches presented in this paper, CL was used as methodology and approach for phraseological and terminographical studies.

CL starts from the quantitative analysis of lexical items occurrence and co-occurrence in authentic language texts, aiming empirical research and/or quantitative/qualitative analysis of language as it is used in real-life communicative contexts. Authentic language analysis provides evidence, based on data, on the functioning and complexity of language in use. Thus, procedures and practices of CL enable linguistic feature observations and lexical unit mutual attraction tendency, recognition of linguistic patterns, and, consequently, analysis of phraseologies (BERBER SARDINHA, 1999; BERBER SARDINHA et al., 2017).

For the researches described in the following sub-section, procedures of *corpora* compilation and standardization were designed and developed according to CL principles and criteria. Furthermore, the analysis of bilingual phraseologies from TV series subtitles was made by WST, in order to save time in the phraseology identification and analysis process.

### 3.2 Step-by-step

Firstly, in order to analyze phraseologies on TV series subtitles, *corpora* of English and Portuguese subtitles were compiled from sites specialized in TV series subtitles, open to free access online. Afterwards, the standardization process was carried out, aiming to facilitate future processes of *data* identification and retrieval.

The following methodological procedures were mapping and extracting the most frequent phraseologies in each English subtitle *corpus*. To do so, WST was used, particularly the WordList, Keyword and Concord tools.

The WordList tool enables quantitative analysis of lexical items occurrence, presenting an increasing or decreasing list of occurring words in the subtitle *corpora*. In the scope of the researches presented in this paper, content words<sup>19</sup> were prioritized for the search for phraseologies and terms, being therefore chosen as their nuclei, as shown in Figure 1.

**Figure 1.** GoT - English subtitle *corpus* word list (partial view).

N	Word	Freq.	%	Texts	%
70	GOOD	593	0,11	60	98,36
71	GO	581	0,11	60	98,36
72	NEVER	574	0,10	60	98,36
73	WHY	567	0,10	60	98,36
74	BACK	529	0,10	60	98,36
75	GET	514	0,09	60	98,36
76	HAS	512	0,09	60	98,36
77	THEIR	509	0,09	60	98,36
78	SEE	508	0,09	60	98,36
79	THINK	507	0,09	60	98,36
80	WERE	501	0,09	60	98,36
81	WELL	499	0,09	60	98,36
82	TAKE	487	0,09	60	98,36
83	HE'S	481	0,09	60	98,36
84	RIGHT	478	0,09	60	98,36
85	OR	474	0,09	60	98,36
86	GRACE	473	0,09	57	93,44
87	MORE	469	0,09	60	98,36
88	TIME	469	0,09	60	98,36
89	LADY	458	0,08	58	95,08
90	I'LL	453	0,08	60	98,36
91	OUT	450	0,08	59	96,72
92	AN	445	0,08	60	98,36
93	BEEN	440	0,08	60	98,36
94	NEED	439	0,08	60	98,36

**Source:** The authors, 2019.

In Figure 1, it is possible to identify some content words, such as **good**, **has**, **need**, **grace**, quantitatively listed in decreasing occurrence order in the Game of Thrones subtitle

<sup>19</sup> In these studies, the focus of research was centered on verbs and nouns, from which the phraseologies were identified.



*corpus*. Thus, the adjective good is more frequent than the verb need, occurring 593 and 439 times, respectively.

Another fundamental procedure for terminographical researches is the compilation of a keywords list, which consists of comparing a study *corpus* word list to a reference *corpus* word list (TAGNIN, 2013), working as a filter to identify words that are candidates for terms in the study *corpus*. According to the author, "this comparison eliminates words with relative similar frequency in the two *corpora* so that words whose frequency is statistically significant remain" (TAGNIN, 2013, p. 218, own translation). In addition, it is advised that the reference *corpus* be adequately balanced, that is, there must be a balance among its constitutive types of text and discursive genres (ALUÍSIO; ALMEIDA, 2006), as well as be at least five times larger than the study *corpus* (BERBER SARDINHA, 2004).

To do so, the WST KeyWords tool was used, in which word lists, obtained by the process presented in the previous paragraphs, were loaded, and, in order to filter candidates for terms, the reference *corpus* used for English was the Corpus of Contemporary American English<sup>20</sup> (COCA), which contains 100 thousand types, pointing out that types are different words which a *corpus* bears in CL.

After identifying the lexical items and candidates for terms, the next procedure was using the Concord tool to access the linguistic context of each item, in order to identify phraseologies that have the lexical items as their nuclei under research. By using Concord, it is possible to select a search-word, and then all search-word occurrences are presented in their linguistic context, that is, flanked by words that precede and succeed it (VIANA; TAGNIN, 2010). The layout of Concord tool data presentation is illustrated in Figure 2.

---

<sup>20</sup> COCA – *Corpus* available at: <http://www2.lael.pucsp.br/corpora/bp/index.htm> . Access on May 18<sup>th</sup>, 2019.

Figure 2. Concord data presentation - layout with go (partial view).

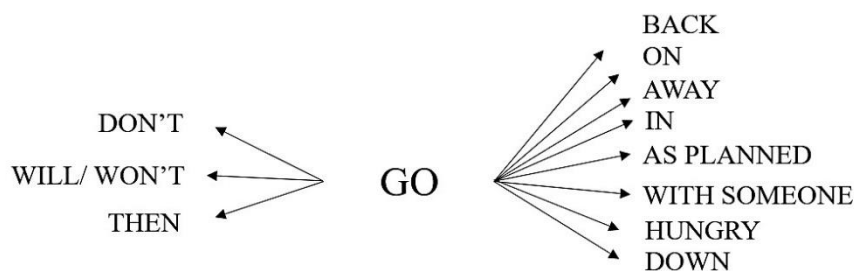
N	Concordance	File	Date
1	513 00:40:00,044 -> 00:40:01,845 If it doesn't go as planned, 514 00:40:01,946 -> 00:40:	GOT_ENG_S06E04.srt	2018/ago/03
2	-> 00:46:10,300 They lose their purpose. They go back to fighting each other. 452 00:46:	GOT_ENG_S04E09.srt	2018/ago/03
3	help. 519 00:48:08,874 -> 00:48:10,208 Then go! 520 00:48:12,011 -> 00:48:14,446 Don't	GOT_ENG_S01E02.srt	2018/ago/02
4	00:48:06,405 -> 00:48:08,840 Stay here, I'll go back to the inn and bring help. 519 00:48:	GOT_ENG_S01E02.srt	2018/ago/02
5	00:40:58,105 -> 00:41:00,305 Please don't go, Jaqen. 611 00:41:03,476 -> 00:41:05,678	GOT_ENG_S02E10.srt	2018/ago/02
6	-> 00:41:29,648 Dog, I command you to go back out there and fight. 572 00:41:33,686	GOT_ENG_S02E09.srt	2018/ago/02
7	:52,744 -> 00:36:55,412 And Highgarden will go to the children 524 00:36:55,446 -> 00:36:	GOT_ENG_S03E06.srt	2018/ago/03
8	-> 00:45:02,674 because when the knife go in 542 00:45:02,725 -> 00:45:06,061 and I	GOT_ENG_S05E05.srt	2018/ago/03
9	554 00:39:47,652 -> 00:39:50,353 - May I go? - As you wish. 555 00:39:56,060 -> 00:39	GOT_ENG_S01E07.srt	2018/ago/02
10	to Joffrey. 523 00:48:29,428 -> 00:48:31,563 Go on- run. 524 00:48:34,667 -> 00:48:	GOT_ENG_S01E02.srt	2018/ago/02
11	made me 540 00:42:36,322 -> 00:42:40,075 go with him down from the Red Keep, into the	GOT_ENG_S05E04.srt	2018/ago/03
12	on- run. 524 00:48:34,667 -> 00:48:36,835 Go! Leave now! 525 00:48:38,938 -> 00:48:	GOT_ENG_S01E02.srt	2018/ago/02
13	. 515 00:36:26,880 -> 00:36:31,016 You may go hungry. You may fall sick. 516 00:36:31,051	GOT_ENG_S03E03.srt	2018/ago/03
14	21,263 -> 00:46:23,847 They'll kill me? If I don't go, they'll kill me anyway. 459 00:46:23,898 ->	GOT_ENG_S04E09.srt	2018/ago/03
15	Mercy! 468 00:40:09,302 -> 00:40:10,840 I'll go! I will! 469 00:40:11,207 -> 00:40:12,672	GOT_ENG_S05E03.srt	2018/ago/03
16	your queen! 472 00:44:37,028 -> 00:44:38,779 Go! 473 00:45:02,887 -> 00:45:04,888 Your	GOT_ENG_S05E09.srt	2018/ago/03
17	00:48:25,424 -> 00:48:29,394 You've got to go. They'll kill you for what you did to Joffrey.	GOT_ENG_S01E02.srt	2018/ago/02
18	-> 00:37:23,628 It is where the warlocks go to squint at dusty books 617 00:37:23,662	GOT_ENG_S02E05.srt	2018/ago/02
19	I did my duty and now she's forcing me to go back? 454 00:37:45,689 -> 00:37:48,423	GOT_ENG_S05E07.srt	2018/ago/03
20	00:30:18,747 -> 00:30:20,915 - Come in. - Go away. 434 00:30:20,949 -> 00:30:23,151	GOT_ENG_S04E01.srt	2018/ago/03
21	here? 491 00:46:17,863 -> 00:46:20,332 - Go away. - Your sister? 492 00:46:24,070 ->	GOT_ENG_S01E02.srt	2018/ago/02
22	-> 00:41:25,124 Do you think Father will let me go down into the crypt? 502 00:41:25,126 ->	GOT_ENG_S05E05.srt	2018/ago/03
23	cave. 413 00:42:33,201 -> 00:42:35,202 We'll go back there. 414 00:42:41,293 -> 00:42:	GOT_ENG_S04E09.srt	2018/ago/03
24	is. 442 00:33:04,929 -> 00:33:08,263 Now, go drink till it feels like you did the right thing.	GOT_ENG_S04E02.srt	2018/ago/03
25	attention. 461 00:29:22,194 -> 00:29:24,896 Go on. Write. 462 00:29:29,151 -> 00:29:	GOT_ENG_S04E03.srt	2018/ago/03

Source: The authors, 2019.

The display of Concord, as shown in Figure 2, enables the identification of patterns, as well as the agglutination tendency of words that co-locate with the search-word (*clusters*), in a faster and more precise way, thus saving time in the qualitative analysis of phraseologies. In addition, Concord presents the files where each occurrence of the search-word takes place, enabling retrieval of their correspondent form in the Portuguese subtitle *corpora*.

Figure 2 presents some occurrences of **to go** in its linguistic context. Consequently, it is possible to identify some phraseologies such as: *to go as planned*, *to go back (there/out there/to)*, *to go in*, *to go on*, *to go with someone*, *to go hungry*, *to go away*, *to go down*. Figure 3 presents an overview of the identified phraseologies:

Figure 3. Partial view of phraseologies with go.



Source: The authors, 2019.

After mapping and analyzing English subtitle *corpora* phraseologies, the *corpora* of Portuguese subtitles were used to identify extracted and analyzed English phraseologies correspondent lexical units. As seen in Figure 2, the Concord tool presents the files containing each search-word occurrences, and thus, the phraseologies. Hence, the files identification was made, and afterwards, the corresponding lexical units, in Portuguese, of each one of the extracted and analyzed phraseologies were identified by using subtitles time markings of each file. After identifying the correspondent lexical units, in English and Portuguese, it was analyzed whether the English phraseologies were also conceived as phraseologies in the Portuguese subtitle *corpora*.

## 4 RESULTS

In this section, we present the results obtained from the SI and PCC phraseological and terminographical researches data analysis, made available in part on VoTec platform, on the entries and example sections. We will list the phraseologies in English, their correspondents in Portuguese and classify them according to the Theoretical Framework discussed in Section 2.

### 4.1 *Bones*<sup>21</sup>

An investigative TV series based on forensic anthropology, with analysis of victims' bones in general.

CLASSIFICATION	ENGLISH PHRASEOLOGIES	PORTUGUESE CORRESPONDENTS	CLASSIFICATION
Noun collocation	a <b>survival technique</b>	<i>uma técnica de sobrevivência</i>	Noun collocation
Noun collocation	<b>identity confusion</b>	<i>confusão de identidade</i>	Noun collocation
Speech formulae	<b>All right</b>	<i>Tudo bem</i>	Speech formulae
Speech formulae	<b>Hands behind your back</b>	<i>Mãos atrás das costas</i>	Speech formulae
Motivated idiom/ colligation of prep./ adjective collocation	he was <b>shot</b> three times <b>at close range</b>	<i>ele foi baleado três vezes à queima-roupa</i>	Frozen idiom/ colligation of prep.
Phrasal verb	hoping to <b>cash in</b>	<i>na esperança de lucrar</i>	-
Verbal regency colligation	But I <b>refuse to</b> perform the autopsy	<i>Mas recuso-me a realizar a autópsia</i>	Verbal regency colligation
Speech formulae	<b>Neither have I</b>	<i>Nem eu</i>	Speech formulae

<sup>21</sup> More details available at: [https://pt.wikipedia.org/wiki/Bones\\_\(s%C3%A9rie\\_de\\_TV\)](https://pt.wikipedia.org/wiki/Bones_(s%C3%A9rie_de_TV)).

REVISTA ELETRÔNICA DE GRADUAÇÃO E PÓS-GRADUAÇÃO EM EDUCAÇÃO REJ/UFG		ITINERARIUS REFLECTIONIS NÚMERO ESPECIAL VOL. 15 Nº2   2019 ISSN 1807-9342	
Verb collocation	You <b>confirm</b> my <b>alibi</b>	<i>Você <b>confirma</b> meu <b>álibi</b></i>	Verb collocation
Verb collocation/ colligation of prep./ restricted collocation	<b>Take</b> these three <b>into custody</b>	<i><b>Tome</b> estes três <b>em custódia</b></i>	Verb collocation/ colligation of prep./ restricted collocation

#### 4.2 Game of Thrones<sup>22</sup>

An epic fantasy TV series depicting various kingdoms and diverse linguistic communities living in war, whose biggest enemies were the White Walkers.

CLASSIFICATION	ENGLISH PHRASEOLOGIES	PORTUGUESE CORRESPONDENTS	CLASSIFICATION
Adjective and noun collocation	She had a <b>Valyrian steel sword</b>	<i>Sua espada era de aço valiriano</i>	Adjective and noun collocation
Speech formulae	Careful, <b>Your Grace</b> .	<i>Cuidado, Vossa Graça/Alteza.</i>	Speech formulae
Frozen idiom/ noun collocation	The <b>Night's Watch</b> is my house now.	<i>A Patrulha da Noite agora é minha Casa</i>	Frozen idiom/ noun collocation
Speech formulae	<b>So they say</b> .	<i>É o que falam.</i>	Speech formulae
Simile	<b>Nothing cuts like Valyrian steel</b> .	<i>Aço valiriano corta sem igual.</i>	Adverbial collocation
Adverbial collocation	If the Night's Watch are <b>truly brothers</b>	<i>Se a Patrulha da Noite é mesmo uma irmandade</i>	Adverbial collocation
Binomial	They'll ring the bells all <b>day and night</b>	<i>Os sinos dobrarão dia e noite</i>	Binomial
Paroemia	<b>Night gathers and now my watch begins</b> .	<i>A noite chega e agora começa a minha vigia.</i>	Paroemia
Motivated idiom	I'm ready to <b>meet the gods</b>	<i>Estou pronta para encontrar os deuses</i>	Motivated idiom
Metaphor	<b>Information is the key</b>	<i>Informação é o segredo.</i>	Metaphor

#### 4.3 Star Trek<sup>23</sup>

Science fiction TV series, whose theme is interstellar travel under the command of Captain Kirk on USS Enterprise spacecraft. The series seeks to convey altruistic ideals through a plot of suspense and adventure.

CLASSIFICATION	ENGLISH PHRASEOLOGIES	PORTUGUESE CORRESPONDENTS	CLASSIFICATION
Noun collocation	equals the <b>speed of light</b>	<i>é igual à velocidade da luz</i>	Noun collocation
Adjective collocation	It's <b>standard protocol</b>	<i>É protocolo padrão</i>	Adjective collocation

<sup>22</sup> More details available at: <<https://www.hbo.com/game-of-thrones>>, Game of Thrones official webpage.

<sup>23</sup> More details available at: <<http://www.startrek.com>>, Star Trek official webpage.



Frozen idiom/ verbal regency colligation	I'm <b>counting on</b> it.	<i>Estou contando com isso.</i>	Frozen idiom/ verbal regency colligation
Phrasal verb	I <b>ended up</b> in the Infirmary	<i>Eu acabei na Enfermaria</i>	-
Speech formulae	It's going to be one <b>hell of an</b> explosion.	<i>Vai ser uma explosão daquelas</i>	-
Speech formulae	<b>What was the point in</b> going to the holodeck?	<i>Qual o sentido em ir para o holodeck?</i>	Speech formulae
Verb collocation	I learned to <b>sing chords</b>	<i>aprendi a entoar mantras</i>	Verb collocation
Simile	He'd used the computer room <b>as if it were an amusement gallery.</b>	<i>Ele usaria/usava o quarto de computador como se fosse uma galeria de diversão.</i>	Simile
Adverbial collocation	It is <b>heavily armoured</b> with an exoskeleton	<i>É fortemente blindado com um exoesqueleto</i>	Adverbial collocation
Verbal regency colligation	Information and details <b>stored in</b> a program	<i>parâmetros gravados em programas de computador</i>	Verbal regency colligation

## 5 FINAL REMARKS

In this paper we aimed to present concepts of phraseology as a discipline and as a Lexicology, Terminology and Semantics object of study. We defined concepts of phraseologies, their subtypes and exemplified them, by Brazilian and Russian scholars for colligation, collocation, phraseme and proverbs. The descriptive contribution of Tagnin (2005) is quite broad, and the theories by Lakoff and Johnson (2003), Dobrovolskij (2012), Barbosa (2012) and Muryn et al. (2013) contribute to the understanding of concepts and cognitive processes that underlie the universe of phraseologies.

The positive aspects of using CL as an approach and methodology in these phraseological and terminographical studies were: (1) analysis of the actual use of English language by native speakers in authentic communicative contexts; (2) undergraduate's access to CL and phraseological/terminographical studies; and (3) training to the production of teaching material for English language classes.

Although they are fictional works, the language used in TV series is real language at use, promoting authenticity and certain representativeness. In the future, this type of study and results can contribute to foreign language teaching, by their use as teaching material in language classrooms.

## REFERENCES





ALUÍSIO, S. M.; ALMEIDA, G. M. de B. O que é e como se constrói um *corpus*: lições aprendidas na compilação de vários *corpora* para pesquisa linguística. **Calidoscópico**, São Leopoldo, v. 4, n. 3, p. 156-178. 2006. Available at <<http://revistas.unisinos.br/index.php/calidoscopio/article/view/6002>>. Access on May 21st, 2019.

ÁLVAREZ, C. A. C. Fraseoparemiología e interculturalidad. In: ALVAREZ, M. L. O. (Org.). **Tendências atuais na pesquisa descritiva e aplicada em fraseologia e paremiologia** – Anais – Vol. 1. Campinas, SP: Pontes Editores, 2012.

ALVAREZ, M. L. O. (Org.). Apresentação: entabulando a conversação... In: SILVA, S. (Org.) **Fraseologia & Cia** – entabulando diálogos reflexivos – 2ª ed. – ampliada, Campinas, SP: Pontes Editores, 2014.

ALVAREZ, M. L. O. (Org.). Dando conta do recado: a motivação metafórica das expressões idiomáticas. In: SILVA, S. (Org.) **Fraseologia & Cia** – entabulando diálogos reflexivos – 2ª ed. – ampliada, Campinas, SP: Pontes Editores, 2014.

BALLY, Ch. **Traité de stylistique française**. 2 vols. Heidelberg, 1909.

BARBOSA, M. A. A fraseologia no percurso gerativo de enunciação de codificação: no sistema, nas normas, no falar concreto. In: **Tendências atuais na pesquisa descritiva e aplicada em fraseologia e paremiologia** – Anais – Vol. 1. Campinas, SP: Pontes Editores, 2012.

BERBER SARDINHA, T. Beginning Portuguese corpus linguistics: exploring a corpus to teach Portuguese as a foreign language. **D.E.L.T.A.**, São Paulo, v. 15, n. 2, p. 289-299. 1999. Available at <<https://revistas.pucsp.br/delta/article/view/40450>>. Access on May 21<sup>st</sup>, 2019.

BERBER SARDINHA, T. Linguística de corpus: histórico e problemática. **D.E.L.T.A.**, São Paulo, v. 16, n. 2, p. 323-367. 2000. Available at <<https://revistas.pucsp.br/delta/article/view/39903>>. Access on May 21<sup>st</sup>, 2019.

BERBER SARDINHA, T. Uso de corpora na formação de tradutores. **D.E.L.T.A.**, São Paulo, v. 19, n. 3, p. 43-70. 2003. Available at <<https://revistas.pucsp.br/delta/article/view/38327>>. Access on May 21<sup>st</sup>, 2019.

BERBER SARDINHA, T. **Linguística de Corpus**. Barueri: Manole, 2004.

BERBER SARDINHA, T. et al. Preparação de material didático para ensino de línguas com base em corpora. **The ESPECIALIST**, São Paulo, v. 38, n. 1. 2017. Available at <<https://revistas.pucsp.br/index.php/esp/article/view/32224>>. Access on March 21st, 2019.

BEVILACQUA, C. R. **A fraseologia jurídico-ambiental**. 1996. 132 f. Dissertação (Mestrado em Letras: Estudos da Linguagem) – Instituto de Letras, Universidade Federal do Rio Grande



do Sul, Porto Alegre. Available at <<http://www.ufrgs.br/termisul/files/file732743.pdf>>. Access on May 11th, 2019.

CHACOTO, L. Fraseoparemiologia e tradutologia. In: **Tendências atuais na pesquisa descritiva e aplicada em fraseologia e paremiologia** – Anais – Vol. 1. Campinas, SP: Pontes Editores, 2012.

DOBROVOL'SKIJ, D.; PIIRAINEN, E. **Figurative Language: Cross-Cultural and Cross-Linguistics Perspectives**. Amsterdam: Elsevier, 2005.

DOBROVOL'SKIJ, D. Phraseology: historical development and theoretical aspects. In: ALVAREZ, M. L. O. (Org.). **Tendências atuais na pesquisa descritiva e aplicada em fraseologia e paremiologia** – Anais – Vol. 1. Campinas, SP: Pontes Editores, 2012.

FLEISCHER, W. **Phraseologie der deutschen Gegenwartssprache**. 2. Aufl: Tübingen, 1997.

FROMM, G. **VoTec: a construção de vocabulários eletrônicos para aprendizes de tradução**. São Paulo, 2007. Tese (Doutorado em Estudos Linguísticos e Literários em Língua Inglesa). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo. Available at <<http://www.teses.usp.br/teses/disponiveis/8/8147/tde-08072008-150855/pt-br.php>>. Access on May 21<sup>st</sup>, 2019.

GARCIA, A. A. F. Compilação e uso de corpora paralelos. In: TAGNIN, S. E. O.; VALE, O. A. (Orgs.) **Avanços da Linguística de Corpus no Brasil**. São Paulo: Humanitas. 2008. p. 117-136.

HALLIDAY, M. A. K. Corpus Studies and Probabilistic Grammar. In: AMMER, K., ALTENBERG, B. (orgs.) **English Corpus Linguistics: Studies in Honour of Jan Svartvik**. Harlow: Longman. 1991. p. 30-43.

KLARE, J. Lexicologia e fraseologia no português moderno. In: **Revista de Filologia Românica**, IV. Madrid: Editorial de la Universidad Complutense, 1986.

LAKOFF, G.; JOHNSON, M. **Metaphors we live by**. Chicago: University of Chicago, 2003.

LEWIS, M. **The Lexical Approach: The State of ELT and a Way Forward**. London, EN.: LTP, 1999.

MURYN, T. ; MEJRI, S. ; PRAŽUCH, W. ; SFAR, I. (Éds.). **La phraséologie entre langues et cultures: structures, fonctionnements, discours**. Frankfurt: Internationaler Verlag der Wissenschaften, 1 edition, 2013.

SCOTT, M. **WordSmith Tools**. Versão 6, 2012. Available at <<https://www.lexically.net/wordsmith/downloads/>>. Access on Sept. 18th, 2018.

TAGNIN, S. E. O. **O jeito que a gente diz: expressões convencionais e idiomáticas**. São Paulo: Disal, 2005.



VIANA, V.; TAGNIN S. E. O. **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, 2010.

TAGNIN, S. E. O. Glossário de Linguística de Corpus. In: TAGNIN, S. E. O.; BEVILACQUA, C. R. (Org.). **Corpora na terminologia**. São Paulo: Hub Editorial, 2013. p. 215-219.